

# Characterizing Realizability in Abstract Argumentation\*

Thomas Linsbichler  
TU Wien  
Austria

Jörg Pührer and Hannes Strass  
Leipzig University  
Germany

## Abstract

Realizability for knowledge representation formalisms studies the following question: Given a semantics and a set of interpretations, is there a knowledge base whose semantics coincides exactly with the given interpretation set? We introduce a general framework for analyzing realizability in abstract dialectical frameworks (ADFs) and various of its subclasses. In particular, the framework applies to Dung argumentation frameworks, SETAFs by Nielsen and Parsons, and bipolar ADFs. We present a uniform characterization method for the admissible, complete, preferred and model/stable semantics. We employ this method to devise an algorithm that decides realizability for the mentioned formalisms and semantics; moreover the algorithm allows for constructing a desired knowledge base whenever one exists. The algorithm is built in a modular way and thus easily extensible to new formalisms and semantics. We have also implemented our approach in answer set programming, and used the implementation to obtain several novel results on the relative expressiveness of the abovementioned formalisms.

## 1 Introduction

The abstract argumentation frameworks (AFs) introduced by Dung (1995) have garnered increasing attention in the recent past. In his seminal paper, Dung showed how an abstract notion of argument (seen as an atomic entity) and the notion of individual attacks between arguments together could reconstruct several established KR formalisms in argumentative terms. Despite the generality of those and many more results in the field that was sparked by that paper, researchers also noticed that the restriction to *individual attacks* is often overly limiting, and devised extensions and generalizations of Dung’s frameworks: directions included generalizing individual attacks to *collective attacks* (Nielsen and Parsons, 2006), leading to so-called SETAFs; others started offering a *support* relation between arguments (Cayrol and Lagasquie-Schiex, 2005), preferences among arguments (Amgoud and Cayrol, 2002; Modgil, 2009), or attacks on attacks into arbitrary depth (Baroni et al., 2011). This is only the tip of an iceberg, for a more comprehensive overview we refer to the work of Brewka, Polberg, and Woltran (2014).

One of the most recent and most comprehensive generalizations of AFs has been presented by Brewka and Woltran (2010) (and later continued by Brewka et al., 2013) in the form of *abstract dialectical frameworks (ADFs)*. These ADFs offer any type of link between arguments: individual attacks (as in AFs), collective attacks (as in SETAFs), and individual and collective support, to name only a few. This generality is achieved through so-called *acceptance conditions* associated to each statement. Roughly, the meaning of relationships between arguments is not fixed in ADFs, but is specified by the user for each argument in the form of Boolean functions (acceptance functions) on the argument’s parents. However, this generality comes with a price: Strass and Wallner (2015) found that the complexity of the associated reasoning problems of ADFs is in general higher than in AFs (one level up in the polynomial hierarchy). Fortunately, the subclass of *bipolar ADFs* (defined by Brewka and Woltran, 2010) is as complex as AFs (for all considered semantics) while still offering a wide range of modeling capacities (Strass and Wallner, 2015). However, there has only been little concerted effort so far to exactly analyze and compare the expressiveness of the abovementioned languages.

This paper is about exactly analyzing means of expression for argumentation formalisms. Instead of motivating expressiveness in natural language and showing examples that some formalisms seem to be able to express but others do not, we tackle the problem in a formal way. We use a precise mathematical definition of expressiveness: a set of interpretations is *realizable* by a formalism under a semantics if and only if there exists a knowledge base of the formalism whose semantics is exactly the given set of interpretations. Studying realizability in AFs has been started by Dunne et al. (2013, 2015), who analyzed realizability for extension-based semantics, that is, interpretations represented by sets where arguments are either accepted (in the extension set) or not accepted (not in the extension set). While their initial work disregarded arguments that are never accepted, there have been continuations where the existence of such “invisible” arguments is ruled out (Baumann et al., 2014; Linsbichler, Spanring, and Woltran, 2015). Dyrkolbotn (2014) began to analyze realizability for labeling-based semantics of AFs, that is, three-valued semantics where arguments can be accepted (mapped to true), rejected (mapped to false) or neither (mapped to unknown). Strass (2015) started to ana-

\*This research has been supported by DFG (project BR 1817/7-1) and FWF (projects I1102 and P25518).

lyze the relative expressiveness of two-valued semantics for ADFs (relative with respect to related formalisms). Most recently, Pührer (2015) presented precise characterizations of realizability for ADFs under several three-valued semantics, namely admissible, grounded, complete, and preferred. The term “precise characterizations” means that he gave necessary and sufficient conditions for an interpretation set to be ADF-realizable under a semantics.

The present paper continues this line of work by lifting it to a much more general setting. We combine the works of Dunne et al. (2015), Pührer (2015), and Strass (2015) into a unifying framework, and at the same time extend them to formalisms and semantics not considered in the respective papers: we treat several formalisms, namely AFs, SETAFs, and (B)ADFs, while the previous works all used different approaches and techniques. This is possible because all of these formalisms can be seen as subclasses of ADFs that are obtained by suitably restricting the acceptance conditions.

Another important feature of our framework is that we uniformly use three-valued interpretations as the underlying model theory. In particular, this means that arguments cannot be “invisible” any more since the underlying vocabulary of arguments is always implicit in each interpretation. Technically, we always assume a fixed underlying vocabulary and consider our results parametric in that vocabulary. In contrast, for example, Dyrkolbotn (2014) presents a construction for realizability that introduces new arguments into the realizing knowledge base; we do not allow that. While sometimes the introduction of new arguments can make sense, for example if new information becomes available about a domain or a debate, it is not sensible in general, as these new arguments would be purely technical with an unclear dialectical meaning. Moreover, it would lead to a different notion of realizability, where most of the realizability problems would be significantly easier, if not trivial.

The paper proceeds as follows. We begin with recalling and introducing the basis and basics of our work – the formalisms we analyze and the methodology with which we analyze them. Next we introduce our general framework for realizability; the major novelty is our consistent use of so-called characterization functions, firstly introduced by Pührer (2015), which we adapt to further semantics. The main workhorse of our approach will be a parametric propagate-and-guess algorithm for deciding whether a given interpretation set is realizable in a formalism under a semantics. We then analyze the relative expressiveness of the considered formalisms, presenting several new results that we obtained using an implementation of our framework. We conclude with a discussion.

## 2 Preliminaries

We make use of standard mathematical concepts like functions and partially ordered sets. For a function  $f : X \rightarrow Y$  we denote the *update of  $f$  with a pair  $(x, y) \in X \times Y$*  by  $f|_y^x : X \rightarrow Y$  with  $z \mapsto y$  if  $z = x$ , and  $z \mapsto f(z)$  otherwise. For a function  $f : X \rightarrow Y$  and  $y \in Y$ , its preimage is  $f^{-1}(y) = \{x \in X \mid f(x) = y\}$ . A *partially ordered set* is a pair  $(S, \sqsubseteq)$  with  $\sqsubseteq$  a partial order on  $S$ . A partially ordered set  $(S, \sqsubseteq)$  is a *complete lattice* if and only if every

$S' \subseteq S$  has both a greatest lower bound ( $\text{glb}$ )  $\bigcap S' \in S$  and a least upper bound ( $\text{lub}$ )  $\bigcup S' \in S$ . A partially ordered set  $(S, \sqsubseteq)$  is a *complete meet-semilattice* iff every non-empty subset  $S' \subseteq S$  has a greatest lower bound  $\bigcap S' \in S$  (the *meet*) and every ascending chain  $C \subseteq S$  has a least upper bound  $\bigcup C \in S$ .

**Three-Valued Interpretations** Let  $A$  be a fixed finite set of statements. An *interpretation* is a mapping  $v : A \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$  that assigns one of the truth values true ( $\mathbf{t}$ ), false ( $\mathbf{f}$ ) or unknown ( $\mathbf{u}$ ) to each statement. An interpretation is *two-valued* if  $v(A) \subseteq \{\mathbf{t}, \mathbf{f}\}$ , that is, the truth value  $\mathbf{u}$  is not assigned. Two-valued interpretations  $v$  can be extended to assign truth values  $v(\varphi) \in \{\mathbf{t}, \mathbf{f}\}$  to propositional formulas  $\varphi$  as usual.

The three truth values are partially ordered according to their information content: we have  $\mathbf{u} <_i \mathbf{t}$  and  $\mathbf{u} <_i \mathbf{f}$  and no other pair in  $<_i$ , which intuitively means that the classical truth values contain more information than the truth value unknown. As usual, we denote by  $\leq_i$  the partial order associated to the strict partial order  $<_i$ . The pair  $(\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}, \leq_i)$  forms a complete meet-semilattice with the information meet operation  $\sqcap_i$ . This meet can intuitively be interpreted as *consensus* and assigns  $\mathbf{t} \sqcap_i \mathbf{t} = \mathbf{t}$ ,  $\mathbf{f} \sqcap_i \mathbf{f} = \mathbf{f}$ , and returns  $\mathbf{u}$  otherwise.

The information ordering  $\leq_i$  extends in a straightforward way to interpretations  $v_1, v_2$  over  $A$  in that  $v_1 \leq_i v_2$  iff  $v_1(a) \leq_i v_2(a)$  for all  $a \in A$ . We say for two interpretations  $v_1, v_2$  that  $v_2$  *extends*  $v_1$  iff  $v_1 \leq_i v_2$ . The set  $\mathcal{V}$  of all interpretations over  $A$  forms a complete meet-semilattice with respect to the information ordering  $\leq_i$ . The consensus meet operation  $\sqcap_i$  of this semilattice is given by  $(v_1 \sqcap_i v_2)(a) = v_1(a) \sqcap_i v_2(a)$  for all  $a \in A$ . The least element of  $(\mathcal{V}, \leq_i)$  is the valuation  $v_{\mathbf{u}} : A \rightarrow \{\mathbf{u}\}$  mapping all statements to unknown – the least informative interpretation. By  $\mathcal{V}_2$  we denote the set of two-valued interpretations; they are the  $\leq_i$ -maximal elements of the meet-semilattice  $(\mathcal{V}, \leq_i)$ . We denote by  $[v]_2$  the set of all two-valued interpretations that extend  $v$ . The elements of  $[v]_2$  form an  $\leq_i$ -antichain with greatest lower bound  $v = \bigcap_i [v]_2$ .

**Abstract Argumentation Formalisms** An *abstract dialectical framework (ADF)* is a tuple  $D = (A, L, C)$  where  $A$  is a set of statements (representing positions one can take or not take in a debate),  $L \subseteq A \times A$  is a set of links (representing dependencies between the positions),  $C = \{C_a\}_{a \in A}$  is a collection of functions  $C_a : 2^{\text{par}(a)} \rightarrow \{\mathbf{t}, \mathbf{f}\}$ , one for each statement  $a \in A$ . The function  $C_a$  is the *acceptance condition of  $a$*  and expresses whether  $a$  can be accepted, given the acceptance status of its parents  $\text{par}(a) = \{b \in S \mid (b, a) \in L\}$ . We usually represent each  $C_a$  by a propositional formula  $\varphi_a$  over  $\text{par}(a)$ . To specify an acceptance condition, then, we take  $C_a(M \cap \text{par}(a)) = \mathbf{t}$  to hold iff  $M$  is a model for  $\varphi_a$ .

Brewka and Woltran (2010) introduced a useful subclass of ADFs: an ADF  $D = (A, L, C)$  is *bipolar* iff all links in  $L$  are supporting or attacking (or both). A link  $(b, a) \in L$  is *supporting in  $D$*  iff for all  $M \subseteq \text{par}(a)$ , we have that  $C_a(M) = \mathbf{t}$  implies  $C_a(M \cup \{b\}) = \mathbf{t}$ . Symmetrically, a link  $(b, a) \in L$  is *attacking in  $D$*  iff for all  $M \subseteq \text{par}(a)$ , we

have that  $C_a(M \cup \{b\}) = \mathbf{t}$  implies  $C_a(M) = \mathbf{t}$ . If a link  $(b, a)$  is both supporting and attacking then  $b$  has no actual influence on  $a$ . (But the link does not violate bipolarity.) We write BADFs as  $D = (A, L^+ \cup L^-, C)$  and mean that  $L^+$  contains all supporting links and  $L^-$  all attacking links.

The semantics of ADFs can be defined using an operator  $\Gamma_D$  over three-valued interpretations (Brewka and Woltran, 2010; Brewka et al., 2013). For an ADF  $D$  and a three-valued interpretation  $v$ , the interpretation  $\Gamma_D(v)$  is given by

$$a \mapsto \bigcap_i \{w(\varphi_a) \mid w \in [v]_2\}$$

That is, for each statement  $a$ , the operator returns the consensus truth value for its acceptance formula  $\varphi_a$ , where the consensus takes into account all possible two-valued interpretations  $w$  that extend the input valuation  $v$ . If this  $v$  is two-valued, we get  $[v]_2 = \{v\}$  and thus  $\Gamma_D(v)(a) = v(\varphi_a)$ .

The standard semantics of ADFs are now defined as follows. For ADF  $D$ , an interpretation  $v : A \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$  is

- *admissible* iff  $v \leq_i \Gamma_D(v)$ ;
- *complete* iff  $\Gamma_D(v) = v$ ;
- *preferred* iff it is  $\leq_i$ -maximal admissible;
- a *two-valued model* iff it is two-valued and  $\Gamma_D(v) = v$ .

We denote the sets of interpretations that are admissible, complete, preferred, and two-valued models by  $\text{adm}(D)$ ,  $\text{com}(D)$ ,  $\text{prf}(D)$  and  $\text{mod}(D)$ , respectively. These definitions are proper generalizations of Dung's notions for AFs: For an AF  $(A, R)$ , where  $R \subseteq A \times A$  is the attack relation, the ADF associated to  $(A, R)$  is  $D_{(A, R)} = (A, R, C)$  with  $C = \{\varphi_a\}_{a \in A}$  and  $\varphi_a = \bigwedge_{b: (b, a) \in R} \neg b$  for all  $a \in A$ . AFs inherit their semantics from the definitions for ADFs (Brewka et al., 2013, Theorems 2 and 4). In particular, an interpretation is *stable* for an AF  $(A, R)$  if and only if it is a two-valued model of  $D_{(A, R)}$ .

A SETAF is a pair  $S = (A, X)$  where  $X \subseteq (2^A \setminus \{\emptyset\}) \times A$  is the (set) attack relation. We define three-valued counterparts of the semantics introduced by Nielsen and Parsons (2006), following the same conventions as in three-valued semantics of AFs (Caminada and Gabbay, 2009) and argumentation formalisms in general. Given a statement  $a \in A$  and an interpretation  $v$  we say that  $a$  is *acceptable* wrt.  $v$  if  $\forall (B, a) \in X \exists a' \in B : v(a') = \mathbf{f}$  and  $a$  is *unacceptable* wrt.  $v$  if  $\exists (B, a) \in X \forall a' \in B : v(a') = \mathbf{t}$ . For an interpretation  $v : A \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$  it holds that

- $v \in \text{adm}(S)$  iff for all  $a \in A$ ,  $a$  is acceptable wrt.  $v$  if  $v(a) = \mathbf{t}$  and  $a$  is unacceptable wrt.  $v$  if  $v(a) = \mathbf{f}$ ;
- $v \in \text{com}(S)$  iff for all  $a \in A$ ,  $a$  is acceptable wrt.  $v$  if  $v(a) = \mathbf{t}$  and  $a$  is unacceptable wrt.  $v$  iff  $v(a) = \mathbf{f}$ ;
- $v \in \text{prf}(S)$  iff  $v$  is  $\leq_i$ -maximal admissible; and
- $v \in \text{mod}(S)$  iff  $v \in \text{adm}(S)$  and  $\nexists a \in A : v(a) = \mathbf{u}$ .

For a SETAF  $S = (A, X)$  the corresponding ADF  $D_S$  has acceptance formula  $\varphi_a = \bigwedge_{(B, a) \in X} \bigvee_{a' \in B} \neg a'$  for each statement  $a \in A$ . (Polberg, 2016)

**Proposition 1.** *For any SETAF  $S = (A, X)$  it holds that  $\sigma(S) = \sigma(D_S)$ , where  $\sigma \in \{\text{adm}, \text{com}, \text{prf}, \text{mod}\}$ .*

*Proof.* Given interpretation  $v$  and statement  $a$ , it holds that  $\Gamma_{D_S}(v)(a) = \mathbf{t}$  iff  $\forall w \in [v]_2 : w(a) = \mathbf{t}$  iff  $\forall (B, a) \in X \exists a' \in B : v(a') = \mathbf{f}$  iff  $a$  is acceptable wrt.  $v$  and  $\Gamma_{D_S}(v)(a) = \mathbf{f}$  iff  $\forall w \in [v]_2 : w(a) = \mathbf{f}$  iff  $\exists (B, a) \in X \forall a' \in B : v(a') = \mathbf{t}$  iff  $a$  is unacceptable wrt.  $v$ . Hence  $\sigma(S) = \sigma(D_S)$  for  $\sigma \in \{\text{adm}, \text{com}, \text{prf}, \text{mod}\}$ .  $\square$

**Realizability** A set  $V \subseteq \mathcal{V}$  of interpretations is *realizable* in a formalism  $\mathcal{F}$  under a semantics  $\sigma$  if and only if there exists a knowledge base  $\text{kb} \in \mathcal{F}$  having exactly  $\sigma(\text{kb}) = V$ . Pührer (2015) characterized realizability for ADFs under various three-valued semantics. We will reuse the central notions for capturing the complete semantics in this work.

**Definition 1 (Pührer 2015).** Let  $V$  be a set of interpretations. A function  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  is a *com-characterization* of  $V$  iff: for each  $v \in \mathcal{V}$  we have  $v \in V$  iff for each  $a \in A$ :

- $v(a) \neq \mathbf{u}$  implies  $f(v_2)(a) = v(a)$  for all  $v_2 \in [v]_2$  and
- $v(a) = \mathbf{u}$  implies  $f(v'_2)(a) = \mathbf{t}$  and  $f(v''_2)(a) = \mathbf{f}$  for some  $v'_2, v''_2 \in [v]_2$ .  $\blacktriangle$

From a function of this kind we can build a corresponding ADF by the following construction. For a function  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$ , we define  $D_f$  as the ADF where the acceptance formula for each statement  $a$  is given by

$$\varphi_a^f = \bigvee_{\substack{w \in \mathcal{V}_2, \\ f(w)(a) = \mathbf{t}}} \phi_w \quad \text{with} \quad \phi_w = \bigwedge_{w(a') = \mathbf{t}} a' \wedge \bigwedge_{w(a') = \mathbf{f}} \neg a'$$

Observe that we have  $v(\phi_w) = \mathbf{t}$  iff  $v = w$  by definition. Intuitively, the acceptance condition  $\varphi_a^f$  is constructed such that  $v$  is a model of  $\varphi_a^f$  if and only if we find  $f(v)(a) = \mathbf{t}$ .

**Proposition 2 (Pührer 2015).** *Let  $V \subseteq \mathcal{V}$  be a set of interpretations. (1) For each ADF  $D$  with  $\text{com}(D) = V$ , there is a com-characterization  $f_D$  for  $V$ ; (2) for each com-characterization  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  for  $V$  we have  $\text{com}(D_f) = V$ .*

The result shows that  $V$  can be realized under complete semantics if and only if there is a *com-characterization* for  $V$ .

### 3 A General Framework for Realizability

The main underlying idea of our framework is that all abstract argumentation formalisms introduced in the previous section can be viewed as subclasses of abstract dialectical frameworks. This is clear for ADFs themselves and for BADFs by definition; for AFs and SETAFs it is fairly easy to see. However, knowing that these formalisms can be recast as ADFs is not everything. To employ this knowledge for realizability, we must be able to precisely characterize the corresponding subclasses in terms of restricting the ADFs' acceptance functions. Alas, this is also possible and paves the way for the framework we present in this section. Most importantly, we will make use of the fact that different formalisms and different semantics can be characterized modularly, that is, independently of each other.

Towards a uniform account of realizability for ADFs under different semantics, we start with a new characterization of realizability for ADFs under admissible semantics that is based on a notion similar in spirit to *com-characterizations*.

**Definition 2.** Let  $V$  be a set of interpretations. A function  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  is an *adm-characterization* of  $V$  iff: for each  $v \in \mathcal{V}$  we have  $v \in V$  iff for every  $a \in A$ :

- $v(a) \neq \mathbf{u}$  implies  $f(v_2)(a) = v(a)$  for all  $v_2 \in [v]_2$ .  $\blacktriangle$

Note that the only difference to Definition 1 is dropping the second condition related to statements with truth value  $\mathbf{u}$ .

**Proposition 3.** Let  $V \subseteq \mathcal{V}$  be a set of interpretations. (1) For each ADF  $D$  such that  $\text{adm}(D) = V$ , there is an *adm-characterization*  $f_D$  for  $V$ ; (2) for each *adm-characterization*  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  for  $V$  we have  $\text{adm}(D_f) = V$ .

*Proof.* (1) We define the function  $f_D : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  as  $f_D(v_2)(a) = v_2(\varphi_a)$  for every  $v_2 \in \mathcal{V}_2$  and  $a \in A$  where  $\varphi_a$  is the acceptance formula of  $a$  in  $D$ . We will show that  $f_D$  is an *adm-characterization* for  $V = \text{adm}(D)$ . Let  $v$  be an interpretation. Consider the case  $v \in \text{adm}(D)$  and  $v(a) \neq \mathbf{u}$  for some  $a \in A$  and some  $v_2 \in [v]_2$ . From  $v \leq_i \Gamma_D(v)$  we get  $v_2(\varphi_a) = v(a)$ . By definition of  $f_D$  it follows that  $f_D(v_2)(a) = v(a)$ . Now assume  $v \notin \text{adm}(D)$  and consequently  $v \not\leq_i \Gamma_D(v)$ . There must be some  $a \in A$  such that  $v(a) \neq \mathbf{u}$  and  $v(a) \neq \Gamma_D(v)(a)$ . Hence, there is some  $v_2 \in [v]_2$  with  $v_2(\varphi_a) \neq v(a)$  and  $f_D(v_2)(a) \neq v(a)$  by definition of  $f_D$ . Thus,  $f_D$  is an *adm-characterization*.

(2) Observe that for every two-valued interpretation  $v_2$  and every  $a \in A$  we have  $f(v_2)(a) = v_2(\varphi_a^f)$ . ( $\subseteq$ ): Let  $v \in \text{adm}(D_f)$  be an interpretation and  $a \in A$  a statement such that  $v(a) \neq \mathbf{u}$ . Let  $v_2$  be a two-valued interpretation with  $v_2 \in [v]_2$ . Since  $v \leq_i \Gamma_{D_f}(v)$  we have  $v(a) = v_2(\varphi_a^f)$ . Therefore, by our observation it must also hold that  $f(v_2)(a) = v(a)$ . Thus, by Definition 2,  $v \in V$ . ( $\supseteq$ ): Consider an interpretation  $v$  such that  $v \notin \text{adm}(D_f)$ . We show that  $v \notin V$ . From  $v \notin \text{adm}(D_f)$  we get  $v \not\leq_i \Gamma_{D_f}(v)$ . There must be some  $a \in A$  such that  $v(a) \neq \mathbf{u}$  and  $v(a) \neq \Gamma_{D_f}(v)(a)$ . Hence, there is some  $v_2 \in [v]_2$  with  $v_2(\varphi_a^f) \neq v(a)$  and consequently  $f(v_2)(a) \neq v(a)$ . Thus, by Definition 2 we have  $v \notin V$ .  $\square$

When listing sets of interpretations in examples, for the sake of readability we represent three-valued interpretations by sequences of truth values, tacitly assuming that the underlying vocabulary is given and has an associated total ordering. For example, for the vocabulary  $A = \{a, b, c\}$  we represent the interpretation  $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}\}$  by the sequence  $\mathbf{tfu}$ .

**Example 1.** Consider the sets  $V_1 = \{\mathbf{uuu}, \mathbf{tff}, \mathbf{ftu}\}$  and  $V_2 = \{\mathbf{tff}, \mathbf{ftu}\}$  of interpretations over  $A = \{a, b, c\}$ . The mapping  $f = \{\mathbf{ttt} \mapsto \mathbf{ftt}, \mathbf{ttf} \mapsto \mathbf{tft}, \mathbf{tft} \mapsto \mathbf{ttt}, \mathbf{tff} \mapsto \mathbf{tff}, \mathbf{ftt} \mapsto \mathbf{ftf}, \mathbf{ftf} \mapsto \mathbf{ftt}, \mathbf{fft} \mapsto \mathbf{ttf}, \mathbf{fff} \mapsto \mathbf{ftf}\}$  is an *adm-characterization* for  $V_1$ . Thus, the ADF  $D_f$  has  $V_1$  as its admissible interpretations. Indeed, the realizing ADF has the following acceptance conditions:

$$\begin{aligned} \varphi_a^f &\equiv (a \wedge b \wedge \neg c) \vee (a \wedge \neg b) \vee (\neg a \wedge \neg b \wedge c) \\ \varphi_b^f &\equiv (a \wedge c) \vee (\neg a \wedge b) \vee (\neg a \wedge \neg b \wedge \neg c) \\ \varphi_c^f &\equiv (a \wedge b) \vee (\neg a \wedge b \wedge \neg c) \vee (\neg b \wedge c) \end{aligned}$$

For  $V_2$  no *adm-characterization* exists because  $\mathbf{uuu} \notin V_2$  but the implication of Definition 2 trivially holds for  $a, b$ , and  $c$ .  $\blacksquare$

We have seen that the construction  $D_f$  for realizing under complete semantics can also be used for realizing a set  $V$  of interpretations under admissible semantics. The only difference is that we here require  $f$  to be an *adm-characterization* instead of a *com-characterization* for  $V$ . Note that admissible semantics can be characterized by properties that are easier to check than existence of an *adm-characterization* (see the work of Pührer, 2015). However, using the same type of characterizations for different semantics allows for a unified approach for checking realizability and constructing a realizing ADF in case one exists.

For realizing under the model semantics, we can likewise present an adjusted version of *com-characterizations*.

**Definition 3.** Let  $V \subseteq \mathcal{V}$  be a set of interpretations. A function  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  is a *mod-characterization* of  $V$  if and only if: (1)  $f$  is defined on  $V$  (that is,  $V \subseteq \mathcal{V}_2$ ) and (2) for each  $v \in \mathcal{V}_2$ , we have  $v \in V$  iff  $f(v) = v$ .  $\blacktriangle$

As we can show, there is a one-to-one correspondence between *mod-characterizations* and ADF realizations.

**Proposition 4.** Let  $V \subseteq \mathcal{V}$  be a set of interpretations. (1) For each ADF  $D$  such that  $\text{mod}(D) = V$ , there is a *mod-characterization*  $f_D$  for  $V$ ; (2) vice versa, for each *mod-characterization*  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  for  $V$  we find  $\text{mod}(D_f) = V$ .

*Proof.* (1) Let  $D$  be an ADF with  $\text{mod}(D) = V$ . It immediately follows that  $V \subseteq \mathcal{V}_2$ . To define  $f_D$  we can use the construction in the proof of Proposition 3. It follows directly that for any  $v \in \mathcal{V}_2$ , we find  $f_D(v) = v$  iff  $v \in V$ . Thus  $f_D$  is a *mod-characterization* for  $V$ .

(2) Let  $V \subseteq \mathcal{V}_2$  and  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  be a *mod-characterization* of  $V$ . For any  $v \in \mathcal{V}_2$  we have:

$$\begin{aligned} v \in V &\iff v = f(v) \\ &\iff \forall a \in A : (v(a) = f(v)(a)) \\ &\iff \forall a \in A : (v(a) = \mathbf{t} \leftrightarrow f(v)(a) = \mathbf{t}) \\ &\iff \forall a \in A : (v(a) = \mathbf{t} \leftrightarrow (\exists w \in \mathcal{V}_2 : f(w)(a) = \mathbf{t} \wedge v = w)) \\ &\iff \forall a \in A : (v(a) = \mathbf{t} \leftrightarrow (\exists w \in \mathcal{V}_2 : f(w)(a) = \mathbf{t} \wedge v(\phi_w) = \mathbf{t})) \\ &\iff \forall a \in A : \left( v(a) = \mathbf{t} \leftrightarrow v \left( \bigvee_{\substack{w \in \mathcal{V}_2, \\ f(w)(a) = \mathbf{t}}} \phi_w \right) = \mathbf{t} \right) \\ &\iff \forall a \in A : v(a) = v \left( \bigvee_{\substack{w \in \mathcal{V}_2, \\ f(w)(a) = \mathbf{t}}} \phi_w \right) \\ &\iff \forall a \in A : v(a) = v(\varphi_a^f) \iff v \in \text{mod}(D_f) \quad \square \end{aligned}$$

A related result was given by Strass (2015, Proposition 10). The characterization we presented here fits into the general framework of this paper and is directly usable for our realizability algorithm. Wrapping up, the next result summarizes how ADF realizability can be captured by different types of characterizations for the semantics we considered so far.

**Theorem 5.** Let  $V \subseteq \mathcal{V}$  be a set of interpretations and consider  $\sigma \in \{adm, com, mod\}$ . There is an ADF  $D$  such that  $\sigma(D) = V$  if and only if there is a  $\sigma$ -characterization for  $V$ .

The preferred semantics of an ADF  $D$  is closely related to its admissible semantics as, by definition, the preferred interpretations of  $D$  are its  $\leq_i$ -maximal admissible interpretations. As a consequence we can also describe preferred realizability in terms of *adm*-characterizations. We use the lattice-theoretic standard notation  $\max_{\leq_i} V$  to select the  $\leq_i$ -maximal elements of a given set  $V$  of interpretations.

**Corollary 6.** Let  $V \subseteq \mathcal{V}$  be a set of interpretations. There is an ADF  $D$  with  $prf(D) = V$  iff there is an *adm*-characterization for some  $V' \subseteq \mathcal{V}$  with  $V \subseteq V'$  and  $\max_{\leq_i} V' = V$ .

Finally, we give a result on the complexity of deciding realizability for the mentioned formalisms and semantics.

**Proposition 7.** Let  $\mathcal{F} \in \{AF, SETAF, BADF, ADF\}$  be a formalism and  $\sigma \in \{adm, com, prf, mod\}$  be a semantics. The decision problem “Given a vocabulary  $A$  and a set  $V \subseteq \mathcal{V}$  of interpretations over  $A$ , is there a  $kb \in \mathcal{F}$  such that  $\sigma(kb) = V$ ?” can be decided in nondeterministic time that is polynomial in the size of  $V$ .<sup>1</sup>

*Proof.* For all considered  $\mathcal{F}$  and  $\sigma$ , computing all  $\sigma$ -interpretations of a given witness  $kb \in \mathcal{F}$  can be done in time that is linear in the size of  $V$ . Comparing the result to  $V$  can also be done in linear time.  $\square$

### 3.1 Deciding Realizability: Algorithm 1

Our main algorithm for deciding realizability is a propagate-and-guess algorithm in the spirit of the DPLL algorithm for deciding propositional satisfiability (Gomes et al., 2008). It is generic with respect to (1) the formalism  $\mathcal{F}$  and (2) the semantics  $\sigma$  for which should be realized. To this end, the propagation part of the algorithm is kept exchangeable and will vary depending on formalism and semantics. Roughly, in the propagation step the algorithm uses the desired set  $V$  of interpretations to derive certain necessary properties of the realizing knowledge base (line 2). This is the essential part of the algorithm: the derivation rules (*propagators*) used there are based on characterizations of realizability with respect to formalism and semantics. Once propagation of properties has reached a fixed point (line 7), the algorithm checks whether the derived information is sufficient to construct a knowledge base. If so, the knowledge base can be constructed and returned (line 9). Otherwise (no more information can be obtained through propagation and there is not enough information to construct a knowledge base yet), the algorithm guesses another assignment for the characterization (line 11) and calls itself recursively.

The main data structure that Algorithm 1 operates on is a set of triples  $(v, a, x)$  consisting of a two-valued interpretation  $v \in \mathcal{V}_2$ , an atom  $a \in A$  and a truth value  $x \in \{t, f\}$ . This data structure is intended to represent the  $\sigma$ -characterizations introduced in Definitions 1 to 3. There,

<sup>1</sup>We assume here that the representation of any  $V$  over  $A$  has size  $\Theta(3^{|A|})$ . There might be specific  $V$  with smaller representations, but we cannot assume any better for the general case.

---

#### Algorithm 1 *realize*( $\mathcal{F}, \sigma, V, F$ )

---

**Input:** • a formalism  $\mathcal{F}$   
• a semantics  $\sigma$  for  $\mathcal{F}$   
• a set  $V$  of interpretations  $v : A \rightarrow \{t, f, u\}$   
• a relation  $F \subseteq \mathcal{V}_2 \times A \times \{t, f\}$ , initially empty

**Output:** a  $kb \in \mathcal{F}$  with  $\sigma(kb) = V$  or “no” if none exists

```

1: repeat
2:   set  $F_\Delta := \bigcup_{p \in P_\sigma^\mathcal{F}} p(V, F) \setminus F$ 
3:   set  $F := F \cup F_\Delta$ 
4:   if  $\exists v \in \mathcal{V}_2, \exists a \in A : \{(v, a, t), (v, a, f)\} \subseteq F$  then
5:     return “no”
6:   end if
7: until  $F_\Delta = \emptyset$ 
8: if  $\forall v \in \mathcal{V}_2, \forall a \in A, \exists x \in \{t, f\} : (v, a, x) \in F$  then
9:   return  $kb_\sigma^\mathcal{F}(F)$ 
10: end if
11: choose  $v \in \mathcal{V}_2, a \in A$  with  $(v, a, t) \notin F, (v, a, f) \notin F$ 
12: if realize( $\mathcal{F}, \sigma, V, F \cup \{(v, a, t)\}$ )  $\neq$  “no” then
13:   return realize( $\mathcal{F}, \sigma, V, F \cup \{(v, a, t)\}$ )
14: else
15:   return realize( $\mathcal{F}, \sigma, V, F \cup \{(v, a, f)\}$ )
16: end if

```

---

a  $\sigma$ -characterization is a function  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$  from two-valued interpretations to two-valued interpretations. However, as the algorithm builds the  $\sigma$ -characterization step by step and there might not even be a  $\sigma$ -characterization in the end (because  $V$  is not realizable), we use a set  $F$  of triples  $(v, a, x)$  to be able to represent both partial and incoherent states of affairs. The  $\sigma$ -characterization candidate induced by  $F$  is partial if we have that for some  $v$  and  $a$ , neither  $(v, a, t) \in F$  nor  $(v, a, f) \in F$ ; likewise, the candidate is incoherent if for some  $v$  and  $a$ , both  $(v, a, t) \in F$  and  $(v, a, f) \in F$ . If  $F$  is neither partial nor incoherent, it gives rise to a unique  $\sigma$ -characterization that can be used to construct the knowledge base realizing the desired set of interpretations. The correspondence to the characterization-function is then such that  $f(v)(a) = x$  iff  $(v, a, x) \in F$ .

In our presentation of the algorithm we focused on its main features, therefore the guessing step (line 11) is completely “blind”. It is possible to use common CSP techniques, such as shaving (removing guessing possibilities that directly lead to inconsistency). Finally, we remark that the algorithm can be extended to enumerate all possible realizations of a given interpretation set – by keeping all choice points in the guessing step and thus exhaustively exploring the whole search space.

In the case where the constructed relation  $F$  becomes functional at some point, the algorithm returns a realizing knowledge base  $kb_\sigma^\mathcal{F}(F)$ . For ADFs, this just means that we denote by  $f$  the  $\sigma$ -characterization represented by  $F$  and set  $kb_\sigma^{\text{ADF}}(F) = D^f$ . For the remaining formalisms we will introduce the respective constructions in later subsections.

The algorithm is parametric in two dimensions, namely with respect to the formalism  $\mathcal{F}$  and with respect to the semantics  $\sigma$ . These two aspects come into the algorithm via

$$\begin{aligned}
p_{adm}^{\infty}(V, F) &= \{(v_2, a, v(a)) \mid v \in V, v_2 \in [v]_2, v(a) \neq \mathbf{u}\} \\
p_{adm}^{\neq}(V, F) &= \{(v_2, a, \neg v(a)) \mid v \in \mathcal{V} \setminus V, v_2 \in [v]_2, \\
&\quad v(a) \neq \mathbf{u}, \forall b \in A \setminus v^{-1}(\mathbf{u}), \forall v'_2 \in [v]_2 : \\
&\quad (a, v_2) \neq (b, v'_2) \rightarrow (v'_2, b, v(b)) \in F\} \\
p_{adm}^{\dagger}(V, F) &= \{(v, a, \mathbf{t}), (v, a, \mathbf{f}) \mid v \in \mathcal{V}_2, a \in A, v_{\mathbf{u}} \notin V\} \\
p_{mod}^{\infty}(V, F) &= \{(v, a, v(a)) \mid v \in V, a \in A\} \\
p_{mod}^{\neq}(V, F) &= \{(v, a, \neg v(a)) \mid v \in \mathcal{V}_2 \setminus V, a \in A, \\
&\quad \forall c \in A \setminus \{a\} : (v, c, v(c)) \in F\} \\
p_{mod}^{\dagger}(V, F) &= \{(v, a, \mathbf{t}), (v, a, \mathbf{f}) \mid v \in \mathcal{V}_2, a \in A, V \not\subseteq \mathcal{V}_2\}
\end{aligned}$$

$$\begin{aligned}
p_{com}^{\infty, \mathbf{u}}(V, F) &= \{(v_2, a, \neg \mathbf{x}) \mid v \in V, v_2 \in [v]_2, v(a) = \mathbf{u}, \\
&\quad \mathbf{x} \in \{\mathbf{t}, \mathbf{f}\}, \forall v'_2 \in [v]_2 : v_2 \neq v'_2 \rightarrow (v'_2, a, \mathbf{x}) \in F\} \\
p_{com}^{\neq, \mathbf{t}\mathbf{f}}(V, F) &= \{(v_2, a, \neg v(a)) \mid v \in \mathcal{V} \setminus V, v_2 \in [v]_2, v(a) \neq \mathbf{u}, \\
&\quad \forall b \in A \setminus v^{-1}(\mathbf{u}), \forall v'_2 \in [v]_2 : (a, v_2) \neq (b, v'_2) \rightarrow (v'_2, b, v(b)) \in F, \\
&\quad \forall b \in v^{-1}(\mathbf{u}), \exists v''_2, v'''_2 \in [v]_2 : (v''_2, b, \mathbf{t}), (v'''_2, b, \mathbf{f}) \in F\} \\
p_{com}^{\neq, \mathbf{u}}(V, F) &= \{(v_2, a, \neg \mathbf{x}) \mid v \in \mathcal{V} \setminus V, v_2 \in [v]_2, v(a) = \mathbf{u}, \\
&\quad \forall b \in A \setminus v^{-1}(\mathbf{u}), \forall v'_2 \in [v]_2 : (v_2, b, v(b)) \in F, \\
&\quad \forall b \in v^{-1}(\mathbf{u}) \setminus \{a\} : \exists v''_2, v'''_2 \in [v]_2 : (v''_2, b, \mathbf{t}), \\
&\quad (v'''_2, b, \mathbf{f}) \in F, \forall v''''_2 \in [v]_2 \setminus \{v_2\} : (v''''_2, b, \mathbf{x}) \in F\}
\end{aligned}$$

Figure 1: *Semantics propagators for the complete* ( $P_{com}^{ADF} = \{p_{com}^{\infty, \mathbf{t}\mathbf{f}}, p_{com}^{\infty, \mathbf{u}}, p_{com}^{\neq, \mathbf{t}\mathbf{f}}, p_{com}^{\neq, \mathbf{u}}\}$  with  $p_{com}^{\infty, \mathbf{t}\mathbf{f}}(V, F) = p_{adm}^{\infty}(V, F)$ ), *admissible* ( $P_{adm}^{ADF} = \{p_{adm}^{\infty}, p_{adm}^{\neq}, p_{adm}^{\dagger}\}$ ), and *model semantics* ( $P_{mod}^{ADF} = \{p_{mod}^{\infty}, p_{mod}^{\neq}, p_{mod}^{\dagger}\}$ ).

so-called *propagators*. A propagator is a formalism-specific or semantics-specific set of derivation rules. Given a set  $V$  of desired interpretations and a partial  $\sigma$ -characterization  $F$ , a propagator  $p$  derives new triples  $(v, a, \mathbf{x})$  that must necessarily be part of any total  $\sigma$ -characterization  $f$  for  $V$  such that  $f$  extends  $F$ . In the following, we present semantics propagators for admissible, complete and two-valued model (in (SET)AF terms stable) semantics, and formalism propagators for BADFs, AFs, and SETAFs.

### 3.2 Semantics Propagators

These propagators (cf. Figure 1) are directly derived from the properties of  $\sigma$ -characterizations presented in Definitions 1 to 3. While the definitions provide exact conditions to check whether a given function is a  $\sigma$ -characterization, the propagators allow us to derive definite values of partial characterizations that are necessary to fulfill the conditions for being a  $\sigma$ -characterization.

For admissible semantics, the condition for a function  $f$  to be an *adm*-characterization of a desired set of interpretations  $V$  (cf. Definition 2) can be split into a condition for desired interpretations  $v \in V$  and two conditions for undesired interpretations  $v \notin V$ . Propagator  $p_{adm}^{\infty}$  derives new triples by considering interpretations  $v \in V$ . Here, for all two-valued interpretations  $v_2$  that extend  $v$ , the value  $f(v_2)$  has to be in accordance with  $v$  on  $v$ 's Boolean part, that is, the algorithm adds  $(v_2, a, v(a))$  whenever  $v(a) \neq \mathbf{u}$ . On the other hand,  $p_{adm}^{\neq}$  derives new triples for  $v \notin V$  in order to ensure that there is a two-valued interpretation  $v_2$  extending  $v$  where  $f(v_2)$  differs from  $v$  on a Boolean value of  $v$ . Note that while  $p_{adm}^{\infty}$  immediately allows us to derive information about  $F$  for each desired interpretation  $v \in V$ , propagator  $p_{adm}^{\neq}$  is much weaker in the sense that it only derives a triple of  $F$  if there is no other way to meet the conditions for an undesired interpretation. Special treatment is required for the interpretation  $v_{\mathbf{u}}$  that maps all statements to  $\mathbf{u}$  and is admissible for every ADF. This is not captured by  $p_{adm}^{\infty}$  and  $p_{adm}^{\neq}$  as these deal only with interpretations that have Boolean mappings. Thus, propagator  $p_{adm}^{\dagger}$  serves to check whether  $v_{\mathbf{u}} \in V$ . If this is not the case, the propagator immediately makes the relation  $F$  incoherent and the algorithm

correctly answers “no”.

For complete semantics and interpretations  $v \in V$ , propagator  $p_{com}^{\infty, \mathbf{t}\mathbf{f}}$  derives triples just like in the admissible case. Propagator  $p_{com}^{\infty, \mathbf{u}}$  deals with statements  $a \in A$  having  $v(a) = \mathbf{u}$  for which there have to be at least two  $v_2, v'_2 \in [v]_2$  having  $f(v_2)(a) = \mathbf{t}$  and  $f(v'_2)(a) = \mathbf{f}$ . Hence  $p_{com}^{\infty, \mathbf{u}}$  derives triple  $(v_2, a, \neg \mathbf{x})$  if for all other  $v'_2 \in [v]_2$  we find a triple  $(v'_2, a, \mathbf{x})$ . For interpretations  $v \notin V$  it must hold that there is some  $a \in A$  such that (i)  $v(a) \neq \mathbf{u}$  and  $f(v_2)(a) \neq v(a)$  for some  $v_2 \in [v]_2$  or (ii)  $v(a) = \mathbf{u}$  but for all  $v_2 \in [v]_2$ ,  $f(v_2)$  assigns the same Boolean truth value  $\mathbf{x}$  to  $a$ . Now if neither (i) nor (ii) can be fulfilled by any statement  $b \in A \setminus \{a\}$  due to the current contents of  $F$ , propagators  $p_{com}^{\neq, \mathbf{t}\mathbf{f}}$  and  $p_{com}^{\neq, \mathbf{u}}$  derive triple  $(v_2, a, \neg v(a))$  for  $v(a) \neq \mathbf{u}$  if needed for  $a$  to fulfill (i) and  $(v_2, a, \neg \mathbf{x})$  for  $v(a) = \mathbf{u}$  if needed for  $a$  to fulfill (ii), respectively.

**Example 2.** Consider the set  $V_3 = \{\mathbf{uuu}, \mathbf{fuu}, \mathbf{uuf}, \mathbf{ftf}\}$ . First, we consider a run of  $\text{realize}(\text{ADF}, \text{adm}, V_3, \emptyset)$ . In the first iteration, propagator  $p_{adm}^{\infty}$  ensures that  $F_{\Delta}$  in line 2 contains  $(\mathbf{fff}, a, \mathbf{f})$ ,  $(\mathbf{ftf}, a, \mathbf{f})$ ,  $(\mathbf{ftf}, c, \mathbf{f})$ , and  $(\mathbf{fff}, c, \mathbf{f})$ . Based on the latter three tuples and  $\mathbf{fuf} \notin V_3$ , propagator  $p_{adm}^{\neq}$  derives  $(\mathbf{fff}, a, \mathbf{t})$  in the second iteration which together with  $(\mathbf{fff}, a, \mathbf{f})$  causes the algorithm to return “no”. Consequently,  $V_3$  is not *adm*-realizable. A run of  $\text{realize}(\text{ADF}, \text{com}, V_3, \emptyset)$  on the other hand returns *com*-characterization  $f$  for  $V_3$  that maps  $\mathbf{ttf}$  to  $\mathbf{tff}$ ,  $\mathbf{ftt}$  to  $\mathbf{fft}$ ,  $\mathbf{ftf}$  and  $\mathbf{fff}$  to  $\mathbf{ftf}$  and all other  $v_2 \in \mathcal{V}_2$  to  $\mathbf{fff}$ . Hence, ADF  $D_f$ , given by the acceptance conditions

$$\begin{aligned}
\varphi_a^f &= a \wedge b \wedge \neg c, & \varphi_c^f &= \neg a \wedge b \wedge c, \\
\varphi_b^f &= (\neg a \wedge b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge \neg c)
\end{aligned}$$

has  $V_3$  as its complete semantics. ■

Finally, for two-valued model semantics, propagator  $p_{mod}^{\infty}$  derives new triples by looking at interpretations  $v \in V$ . For those, we must find  $f(v) = v$  in each *mod*-characterization  $f$  by definition. Thus the algorithm adds  $(v, a, v(a))$  for each  $a \in A$  to the partial characterization  $F$ . Propagator  $p_{mod}^{\neq}$  looks at interpretations  $v \in \mathcal{V}_2 \setminus V$ , for which it must hold that  $f(v) \neq v$ . Thus there must be a statement  $a \in A$  with  $v(a) \neq f(v)(a)$ , which is exactly what this propagator

---

**Algorithm 2** *realizePrf*( $\mathcal{F}, V$ )

---

**Input:** • a formalism  $\mathcal{F}$ • a set  $V$  of interpretations  $v : A \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ **Output:** Return some  $\text{kb} \in \mathcal{F}$  with  $\text{prf}(\text{kb}) = V$  if one exists or “no” otherwise.

```
1: if  $\max_{\leq_i} V \neq V$  then
2:   return “no”
3: end if
4: set  $V^< := \{v \in \mathcal{V} \mid \exists v' \in V : v <_i v'\}$ 
5: set  $X := \emptyset$ 
6: repeat
7:   choose  $V' \subseteq V^<$  with  $V' \notin X$ 
8:   set  $X := X \cup \{V'\}$ 
9:   set  $V^{adm} := V \cup V'$ 
10:  if realize( $\mathcal{F}, adm, V^{adm}, \emptyset$ )  $\neq$  “no” then
11:    return realize( $\mathcal{F}, adm, V^{adm}, \emptyset$ )
12:  end if
13: until  $\forall V' \subseteq V^< : V' \in X$ 
14: return “no”
```

---

derives whenever it is clear that there is only one statement candidate left. This, in turn, is the case whenever all  $b \in A$  with the opposite truth value  $\neg v(a)$  and all  $c \in A$  with  $c \neq a$  cannot coherently become the necessary witness any more. The propagator  $p_{mod}^{\sharp}$  checks whether  $V \subseteq \mathcal{V}_2$ , that is, the desired set of interpretations consists entirely of two-valued interpretations. In that case this propagator makes the relation  $F$  incoherent, following a similar strategy as  $p_{adm}^{\sharp}$ .

**Preferred Semantics** Realizing a given set of interpretations  $V$  under preferred semantics requires special treatment. We do not have a  $\sigma$ -characterization function for  $\sigma = \text{prf}$  at hand to directly check realizability of  $V$  but have to find some  $V' \subseteq \{v \in \mathcal{V} \mid \exists v' \in V : v <_i v'\}$  such that  $V \cup V'$  is realizable under admissible semantics (cf. Corollary 6). Algorithm 2 implements this idea by guessing such a  $V'$  (line 7) and then using Algorithm 1 to try to realize  $V \cup V'$  under admissible semantics (line 11). If *realize* returns a knowledge base  $\text{kb}$  realizing  $V \cup V'$  under *adm* we can directly use  $\text{kb}$  as solution of *realizePrf* since it holds that  $\text{prf}(\text{kb}) = V$ , given that  $V$  is an  $\leq_i$ -antichain (line 2).

### 3.3 Formalism Propagators

When constructing an ADF realizing a given set  $V$  of interpretations under a semantics  $\sigma$ , the function  $\text{kb}_{\sigma}^{\text{ADF}}(F)$  makes use of the  $\sigma$ -characterization given by  $F$  in the following way:  $v$  is a model of the acceptance condition  $\varphi_a$  if and only if we find  $(v, a, \mathbf{t}) \in F$ . Now as bipolar ADFs, SETAFs and AFs are all subclasses of ADFs by restricting the acceptance conditions of statements, these restrictions also carry over to the  $\sigma$ -characterizations. The propagators defined below use structural knowledge on the form of acceptance conditions of the respective formalisms to reduce the search space or to induce incoherence of  $F$  whenever  $V$  is not realizable.

**Bipolar ADFs** For bipolar ADFs, we use the fact that each of their links must have at least one polarity, that is, must

be supporting or attacking. Therefore, if a link is not supporting, it must be attacking, and vice versa. For canonical realization, we obtain the polarities of links, i.e. the sets  $L^+$  and  $L^-$ , as defined in Figure 2.

**AFs** To explain the AF propagators, we first need some more definitions. On the two classical truth values, we define the truth ordering  $\mathbf{f} <_t \mathbf{t}$ , whence the operations  $\sqcup_t$  and  $\sqcap_t$  with  $\mathbf{f} \sqcup_t \mathbf{t} = \mathbf{t}$  and  $\mathbf{f} \sqcap_t \mathbf{t} = \mathbf{f}$  result. These operations can be lifted pointwise to two-valued interpretations as usual, that is,  $(v_1 \sqcup_t v_2)(a) = v_1(a) \sqcup_t v_2(a)$  and  $(v_1 \sqcap_t v_2)(a) = v_1(a) \sqcap_t v_2(a)$ . Again, the reflexive version of  $<_t$  is denoted by  $\leq_t$ . The pair  $(\mathcal{V}_2, \leq_t)$  of two-valued interpretations ordered by the truth ordering forms a complete lattice with glb  $\sqcap_t$  and lub  $\sqcup_t$ . This complete lattice has the least element  $v_{\mathbf{f}} : A \rightarrow \{\mathbf{f}\}$ , the interpretation mapping all statements to false, and the greatest element  $v_{\mathbf{t}} : A \rightarrow \{\mathbf{t}\}$  mapping all statements to true, respectively.

Acceptance conditions of AF-based ADFs have the form of conjunctions of negative literals. In the complete lattice  $(\mathcal{V}_2, \leq_t)$ , the model sets of AF acceptance conditions correspond to the lattice-theoretic concept of an *ideal*, a subset of  $\mathcal{V}_2$  that is downward-closed with respect to  $\leq_t$  and upward-closed with respect to  $\sqcup_t$ . The propagator directly implements these closure properties: application of  $p^{\text{AF}}$  ensures that when a  $\sigma$ -characterization  $F$  that is neither incoherent nor partial is found in line 8 of Algorithm 1, then there is, for each  $a \in A$ , an interpretation  $v_a$  such that  $(v_a, a, \mathbf{t}) \in F$  and  $v \leq_t v_a$  for each  $(v, a, \mathbf{t}) \in F$ . Hence  $v_a$  is crucial for the acceptance condition, or in AF terms the attacks, of  $a$  and we can define  $\text{kb}_{\sigma}^{\text{AF}}(F) = (A, \{(b, a) \mid a, b \in A, v_a(b) = \mathbf{f}\})$ .

**SETAFs** The propagator for SETAFs,  $p^{\text{SETAF}}$ , is a weaker version of that of AFs, since we cannot presume upward-closure with respect to  $\sqcup_t$ . In SETAF-based ADFs the acceptance formula is in *conjunctive normal form* containing only negative literals. By a transformation preserving logical equivalence we obtain an acceptance condition in *disjunctive normal form*, again with only negative literals; in other words, a *disjunction* of AF acceptance formulas. Thus, the model set of a SETAF acceptance condition is not necessarily an ideal, but a union of ideals. For the canonical realization we can make use of the fact that, for each  $a \in A$ , the set  $V_a^{\mathbf{t}} = \{v \in \mathcal{V}_2 \mid (v, a, \mathbf{t}) \in F\}$  is downward-closed with respect to  $\leq_t$ , hence the set of models of  $\bigvee_{v \in \max_{\leq_t} V^{\mathbf{t}}} \bigwedge_{v(b)=\mathbf{f}} \neg b$  is exactly  $V_a^{\mathbf{t}}$ . The clauses of its corresponding CNF-formula exactly coincide with the sets of arguments attacking  $a$  in  $\text{kb}_{\sigma}^{\text{SETAF}}(F)$ .

### 3.4 Correctness

For a lack of space, we could not include a formal proof of soundness and completeness of Algorithm 1, but rather present arguments for termination and correctness.

**Termination** With each recursive call, the set  $F$  can never decrease in size, as the only changes to  $F$  are adding the results of propagation in line 3 and adding the guesses in line 11. Also within the until-loop, the set  $F$  can never decrease in size; furthermore there is only an overall finite number of triples that can be added to  $F$ . Thus at some point

$$\begin{aligned}
p^{\text{SETAF}}(V, F) &= \{(v_f, a, \mathbf{t}) \mid a \in A\} \cup \{(w, a, \mathbf{t}) \mid (v, a, \mathbf{t}) \in F, w \in \mathcal{V}_2, w <_t v\} \cup \{(w, a, \mathbf{f}) \mid (v, a, \mathbf{f}) \in F, w \in \mathcal{V}_2, v <_t w\} \\
p^{\text{AF}}(V, F) &= p^{\text{SETAF}}(V, F) \cup \{(v_1 \sqcup_t v_2, a, \mathbf{t}) \mid (v_1, a, \mathbf{t}) \in F, (v_2, a, \mathbf{t}) \in F\} & L^+ &= \{(b, a) \mid (v, a, \mathbf{f}) \in F, v(b) = \mathbf{f}, (v|_t^b, a, \mathbf{t}) \in F\} \\
p^{\text{BADF}}(V, F) &= \{(v|_t^b, a, \mathbf{x}) \mid (v, a, \mathbf{x}) \in F, (w, a, \neg \mathbf{x}) \in F, w(b) = \mathbf{f}, (w|_t^b, a, \mathbf{x}) \in F\} & L^- &= \{(b, a) \mid (v, a, \mathbf{t}) \in F, v(b) = \mathbf{f}, (v|_t^b, a, \mathbf{f}) \in F\}
\end{aligned}$$

Figure 2: Formalism propagators. For formalism  $\mathcal{F} \in \{\text{AF}, \text{SETAF}, \text{BADF}\}$  and any  $\sigma \in \{\text{adm}, \text{com}, \text{prf}, \text{mod}\}$ , we set the respective propagator for  $\mathcal{F}$  to  $P_\sigma^\mathcal{F} = P_\sigma^{\text{ADFF}} \cup \{p^\mathcal{F}\}$  with  $p^\mathcal{F}$  as defined above.  $L^+$  and  $L^-$  define link polarities for  $kb_\sigma^{\text{BADF}}$ .

we must have  $F_\Delta = \emptyset$  and leave the until-loop. Since  $F$  always increases in size, at some point it must either become functional or incoherent, whence the algorithm terminates.

**Soundness** If the algorithm returns a realizing knowledge base  $kb_\sigma^\mathcal{F}(F)$ , then according to the condition in line 8 the relation  $F$  induced a total function  $f : \mathcal{V}_2 \rightarrow \mathcal{V}_2$ . In particular, because the until-loop must have been run through at least once, there was at least one propagation step (line 2). Since the propagators are defined such that they enforce everything that must hold in a  $\sigma$ -characterization, we conclude that the induced function  $f$  indeed is a  $\sigma$ -characterization for  $V$ . By construction, we consequently find that  $\sigma(kb_\sigma^\mathcal{F}(F)) = V$ .

**Completeness** If the algorithm answers “no”, then the execution reached line 5. Thus, for the constructed set  $F$ , there must have been an interpretation  $v \in \mathcal{V}_2$  and a statement  $a \in A$  such that  $\{(v, a, \mathbf{t}), (v, a, \mathbf{f})\} \subseteq F$ , that is,  $F$  is incoherent. Since  $F$  is initially empty, the only way it could get incoherent is in the propagation step in line 2. (The guessing step cannot create incoherence, since exactly one truth value is guessed for  $v$  and  $a$ .) However, the propagators are defined such that they infer only assignments (triples) that are necessary for the given  $F$ . Consequently, the given interpretation set  $V$  is such that either there is no realization within the ADF fragment corresponding to formalism  $\mathcal{F}$  (that is, the formalism propagator derived the incoherence) or there is no  $\sigma$ -characterization for  $V$  with respect to general ADFs (that is, the semantics propagator derived the incoherence). In any case,  $V$  is not  $\sigma$ -realizable for  $\mathcal{F}$ .

## 4 Implementation

As Algorithm 1 is based on propagation, guessing, and checking it is perfectly suited for an implementation using answer set programming (ASP) (Niemelä, 1999; Marek and Truszczyński, 1999) as this allows for exploiting conflict learning strategies and heuristics of modern ASP solvers. Thus, we developed ASP encodings in the Gringo language (Gebser et al., 2012) for our approach. Similar as the algorithm, our declarative encodings are modular, consisting of a main part responsible for constructing set  $F$  and separate encodings for the individual propagators. If one wants, e.g., to compute an AF realization under admissible semantics for a set  $V$  of interpretations, an input program encoding  $V$  is joined with the main encoding, the propagator encoding for admissible semantics as well as the propagator encoding for AFs. Every answer set of such a program encodes a respective characterization function. Our ASP encoding for preferred semantics is based on the admissible encoding and guesses further interpretations following the

essential idea of Algorithm 2. For constructing a knowledge base with the desired semantics, we also provide two ASP encodings that transform the output to an ADF in the syntax of the DIAMOND tool (Ellmauthaler and Strass, 2014), respectively an AF in ASPARTIX syntax (Egly, Gaggl, and Woltran, 2010; Gaggl et al., 2015). Both argumentation tools are based on ASP themselves. The encodings for all the semantics and formalisms we covered in the paper can be downloaded from <http://www.dbai.tuwien.ac.at/research/project/adf/unreal/>. A selection of them is depicted in Figure 3 on the next page.

## 5 Expressiveness Results

In this section we briefly present some results that we have obtained using our implementation. We first introduce some necessary notation to describe the relative expressiveness of knowledge representation formalisms (Gogic et al., 1995; Strass, 2015). For formalisms  $\mathcal{F}_1$  and  $\mathcal{F}_2$  with semantics  $\sigma_1$  and  $\sigma_2$ , we say that  $\mathcal{F}_2$  under  $\sigma_2$  is at least as expressive as  $\mathcal{F}_1$  under  $\sigma_1$  and write  $\mathcal{F}_1^{\sigma_1} \leq_e \mathcal{F}_2^{\sigma_2}$  if and only if  $\Sigma_{\mathcal{F}_1}^{\sigma_1} \subseteq \Sigma_{\mathcal{F}_2}^{\sigma_2}$ , where  $\Sigma_{\mathcal{F}}^\sigma = \{\sigma(kb) \mid kb \in \mathcal{F}\}$  is the *signature of  $\mathcal{F}$  under  $\sigma$* . As usual, we define  $\mathcal{F}_1 <_e \mathcal{F}_2$  iff  $\mathcal{F}_1 \leq_e \mathcal{F}_2$  and  $\mathcal{F}_2 \not\leq_e \mathcal{F}_1$ .

We now start by considering the signatures of AFs, SETAFs and (B)ADFs for the unary vocabulary  $\{a\}$ :

$$\begin{aligned}
\Sigma_{\text{AF}}^{\text{adm}} &= \Sigma_{\text{SETAF}}^{\text{adm}} = \{\{\mathbf{u}\}, \{\mathbf{u}, \mathbf{t}\}\} \\
\Sigma_{\text{AF}}^{\text{com}} &= \Sigma_{\text{SETAF}}^{\text{com}} = \{\{\mathbf{u}\}, \{\mathbf{t}\}\} \\
\Sigma_{\text{AF}}^{\text{prf}} &= \Sigma_{\text{SETAF}}^{\text{prf}} = \{\{\mathbf{u}\}, \{\mathbf{t}\}\} \\
\Sigma_{\text{AF}}^{\text{mod}} &= \Sigma_{\text{SETAF}}^{\text{mod}} = \{\emptyset, \{\mathbf{t}\}\} \\
\Sigma_{\text{ADF}}^{\text{adm}} &= \Sigma_{\text{BADF}}^{\text{adm}} = \Sigma_{\text{AF}}^{\text{adm}} \cup \{\{\mathbf{u}, \mathbf{f}\}, \{\mathbf{u}, \mathbf{t}, \mathbf{f}\}\} \\
\Sigma_{\text{ADF}}^{\text{com}} &= \Sigma_{\text{BADF}}^{\text{com}} = \Sigma_{\text{AF}}^{\text{com}} \cup \{\{\mathbf{f}\}, \{\mathbf{u}, \mathbf{t}, \mathbf{f}\}\} \\
\Sigma_{\text{ADF}}^{\text{prf}} &= \Sigma_{\text{BADF}}^{\text{prf}} = \Sigma_{\text{AF}}^{\text{prf}} \cup \{\{\mathbf{f}\}, \{\mathbf{t}, \mathbf{f}\}\} \\
\Sigma_{\text{ADF}}^{\text{mod}} &= \Sigma_{\text{BADF}}^{\text{mod}} = \Sigma_{\text{AF}}^{\text{mod}} \cup \{\{\mathbf{f}\}, \{\mathbf{t}, \mathbf{f}\}\}
\end{aligned}$$

The following result shows that the expressiveness of the formalisms under consideration is in line with the amount of restrictions they impose on acceptance formulas.

**Theorem 8.** For any  $\sigma \in \{\text{adm}, \text{com}, \text{prf}, \text{mod}\}$ :

1.  $\text{AF}^\sigma <_e \text{SETAF}^\sigma$ .
2.  $\text{SETAF}^\sigma <_e \text{BADF}^\sigma$ .
3.  $\text{BADF}^\sigma <_e \text{ADF}^\sigma$ .

*Proof.* (1)  $\text{AF}^\sigma \leq_e \text{SETAF}^\sigma$  is clear (by modeling individual attacks via singletons). For  $\text{SETAF}^\sigma \not\leq_e \text{AF}^\sigma$  the witnessing model sets over vocabulary  $A = \{a, b, c\}$  are  $\{\mathbf{uuu}, \mathbf{ttf}, \mathbf{tft}, \mathbf{ftt}\} \in \Sigma_{\text{SETAF}}^\sigma \setminus \Sigma_{\text{AF}}^\sigma$  and  $\{\mathbf{ttf}, \mathbf{tft}, \mathbf{ftt}\} \in$



---

Main Encoding

---

```

1 %
2 %
3 cterm(A, t(A)) :- s(A).
4 cterm(A, f(A)) :- s(A).
5
6 %
7 %
8 %
9 %
10 int(nil).
11 int((AS, I)) :- s(A), cterm(A, AS),
12               int(I), smaller(A, I).
13 smaller(A, nil) :- s(A).
14 smaller(A, (H, I)) :- s(A), cterm(T, H), A < T, int(I).
15
16 %
17 member(T, (T, I)) :- int((T, I)).
18 member(T, (X, I)) :- int((X, I)), member(T, I).
19
20 %
21 int2(I) :- int(I), not hasU(I).
22 hasU(I) :- hasU(I, A).
23 hasU(I, A) :- int(I), s(A), not member(t(A), I),
24               not member(f(A), I).
25
26 %
27 ileq(I, J) :- int(I), int(J), not nileq(I, J).
28 nileq(I, J) :- int(I), int(J), member(T, I),
29               not member(T, J).
30
31 %
32 1 { ch(A, I, t); ch(A, I, f) } 1 :- s(A), int2(I).

```

---

Two-Valued Model Encoding

---

```

1 %
2 :- in(I), not int2(I).
3
4 %
5 ch(A, I, t) :- int2(I), in(I), s(A), member(t(A), I).
6 ch(A, I, f) :- int2(I), in(I), s(A), member(f(A), I).
7
8 %
9 ch(A, I, t) :- int2(I), not in(I), member(f(A), I),
10               ch(B, I, t) : s(B), member(t(B), I);
11               ch(C, I, f) : s(C), member(f(C), I), C != A.
12 ch(A, I, f) :- int2(I), not in(I), member(t(A), I),
13               ch(B, I, f) : s(B), member(f(B), I);
14               ch(C, I, t) : s(C), member(t(C), I), C != A.

```

---

BADF Encoding

---

```

1 %
2 %
3 att(B, A) :- ch(A, I, t), ch(A, J, f), diffFT(I, J, B).
4
5 %
6 %
7 sup(B, A) :- ch(A, I, f), ch(A, J, t), diffFT(I, J, B).
8
9 %
10 diffFT(I, J, A) :- int2(I), int2(J), member(f(A), I, B),
11                   member(t(A), J, B).
12 member(T, (T, I), I) :- int((T, I)).
13 member(T, (X, I), (X, B)) :- int((X, I)),
14                               member(T, I, B), X != T.
15
16 %
17 %
18 ch(A, J, f) :- att(B, A), ch(A, I, f), diffFT(I, J, B).
19 ch(A, J, t) :- sup(B, A), ch(A, I, t), diffFT(I, J, B).

```

Figure 3: Selected ASP encodings in clingo 4 syntax. The main encoding implements Algorithm 1, the remaining encodings implement the two-valued model semantics propagator, and the BADF formalism propagator, respectively.

$\Sigma_{\text{SETAF}}^\tau \setminus \Sigma_{\text{AF}}^\tau$  with  $\sigma \in \{\text{adm}, \text{com}\}$  and  $\tau \in \{\text{prf}, \text{mod}\}$ . By each pair of arguments of  $A$  being  $\mathbf{t}$  in at least one model, a realizing AF cannot feature any attack, immediately giving rise to the model  $\mathbf{ttt}$ . The respective realizing SETAF is given by the attack relation  $R = \{(\{a, b\}, c), (\{a, c\}, b), (\{b, c\}, a)\}$ .

(2) It is clear that  $\text{SETAF}^\sigma \leq_e \text{BADF}^\sigma$  holds (all parents are always attacking). For  $\text{BADF}^\sigma \not\leq_e \text{SETAF}^\sigma$  the respective counterexamples can be read off the signatures above: for  $\sigma \in \{\text{adm}, \text{com}\}$  we find  $\{\mathbf{u}, \mathbf{t}, \mathbf{f}\} \in \Sigma_{\text{BADF}}^\sigma \setminus \Sigma_{\text{SETAF}}^\sigma$  and for  $\tau \in \{\text{prf}, \text{mod}\}$  we find  $\{\mathbf{t}, \mathbf{f}\} \in \Sigma_{\text{BADF}}^\tau \setminus \Sigma_{\text{SETAF}}^\tau$ .

(3) For  $\sigma = \text{mod}$  the result is known (Strass, 2015, Theorem 14); for the remaining semantics the model sets witnessing  $\text{ADF}^\sigma \not\leq_e \text{BADF}^\sigma$  over vocabulary  $A = \{a, b\}$  are

$$\begin{aligned}
\{\mathbf{uu}, \mathbf{tu}, \mathbf{tt}, \mathbf{tf}, \mathbf{fu}\} &\in \Sigma_{\text{ADF}}^{\text{adm}} \setminus \Sigma_{\text{BADF}}^{\text{adm}} \\
\{\mathbf{uu}, \mathbf{tu}, \mathbf{tt}, \mathbf{tf}, \mathbf{fu}\} &\in \Sigma_{\text{ADF}}^{\text{com}} \setminus \Sigma_{\text{BADF}}^{\text{com}} \\
\{\mathbf{tt}, \mathbf{tf}, \mathbf{fu}\} &\in \Sigma_{\text{ADF}}^{\text{prf}} \setminus \Sigma_{\text{BADF}}^{\text{prf}}
\end{aligned}$$

A witnessing ADF is given by  $\varphi_a = a$  and  $\varphi_b = a \leftrightarrow b$ .  $\square$

Theorem 8 is concerned with the relative expressiveness of the formalisms under consideration, given a certain semantics. Considering different semantics we find that for all formalisms the signatures become incomparable:

**Proposition 9.**  $\mathcal{F}_1^{\sigma_1} \not\leq_e \mathcal{F}_2^{\sigma_2}$  and  $\mathcal{F}_2^{\sigma_2} \not\leq_e \mathcal{F}_1^{\sigma_1}$  for all formalisms  $\mathcal{F}_1, \mathcal{F}_2 \in \{\text{AF}, \text{SETAF}, \text{BADF}, \text{ADF}\}$  and all semantics  $\sigma_1, \sigma_2 \in \{\text{adm}, \text{com}, \text{prf}, \text{mod}\}$  with  $\sigma_1 \neq \sigma_2$ .

*Proof.* First, the result for *adm* and *com* follows by  $\{\mathbf{u}, \mathbf{t}\} \in \Sigma_{\text{AF}}^{\text{adm}}$ , but  $\{\mathbf{u}, \mathbf{t}\} \notin \Sigma_{\text{ADF}}^{\text{com}}$  and  $\{\mathbf{t}\} \in \Sigma_{\text{AF}}^{\text{com}}$ , but  $\{\mathbf{t}\} \notin \Sigma_{\text{ADF}}^{\text{adm}}$ . Moreover, taking into account that the set of preferred interpretations (resp. two-valued models) always forms a  $\leq_i$ -antichain while the set of admissible (resp. complete) interpretations never does, the result follows for  $\sigma_1 \in \{\text{adm}, \text{com}\}$  and  $\sigma_2 \in \{\text{prf}, \text{mod}\}$ . Finally, since a  $\text{kb} \in \mathcal{F}$  may not have any two-valued models and a preferred interpretation is not necessarily two-valued, the result for *prf* and *mod* follows.  $\square$

Disregarding the possibility of realizing the empty set of interpretations under the two-valued model semantics, we

obtain the following relation for ADFs.

**Proposition 10.**  $(\Sigma_{ADF}^{mod} \setminus \{\emptyset\}) \subseteq \Sigma_{ADF}^{prf}$ .

*Proof.* Consider some  $V \in \Sigma_{ADF}^{mod}$  with  $V \neq \emptyset$ . Clearly  $V \subseteq \mathcal{V}_2$  and by Proposition 4 there is a *mod*-characterization  $f: \mathcal{V}_2 \rightarrow \mathcal{V}_2$  for  $V$ , that is,  $f(v) = v$  iff  $v \in V$ . Define  $f': \mathcal{V}_2 \rightarrow \mathcal{V}_2$  such that  $f'(v) = f(v) = v$  for all  $v \in V$  and  $f'(v)(a) = \neg v(a)$  for all  $v \in \mathcal{V} \setminus V$  and  $a \in A$ . Now it holds that  $f'$  is an *adm*-characterization of  $V' = \{v \in \mathcal{V} \mid \forall v_2 \in [v]_2 : v_2 \in V\} \cup \{v_u\}$ . Since  $\max_{\leq_i} V' = V$  we get that the ADF  $D$  with acceptance formula  $\varphi_a^{f'}$  for each  $a \in A$  has  $prf(D) = V$  whence  $V \in \Sigma_{ADF}^{prf}$ .  $\square$

In contrast, this relation does not hold for AFs, which was shown for extension-based semantics by Linsbichler, Spanring, and Woltran (2015) (Theorem 5) and immediately follows for the three-valued case.

## 6 Discussion

We presented a framework for realizability in which AFs, SETAFs, BADFs and general ADFs can be treated in a uniform way. The centerpiece of our approach is an algorithm for deciding realizability of a given interpretation-set in a formalism under a semantics. The algorithm makes use of so-called propagators, by which it can be adapted to the different formalisms and semantics. We also presented an implementation of our framework in answer set programming and several novel expressiveness results that we obtained using our implementation. In related work, Polberg (2016) studies a wide range of abstract argumentation formalisms, in particular their relationship with ADFs. This can be the basis for including further formalisms into our realizability framework: all that remains to do is figuring out suitable ADF fragments and developing propagators for them, just like we did exemplarily for Nielsen and Parsons's SETAFs. For further future work, we could also streamline existing propagators such that they do not only derive absolutely necessary assignments, but also logically weaker conclusions, such as disjunctions of (non-)assignments.

## References

- Amgoud, L., and Cayrol, C. 2002. A reasoning model based on the production of acceptable arguments. *Ann. Math. Artif. Intell.* 34(1–3):197–215.
- Baroni, P.; Cerutti, F.; Giacomin, M.; and Guida, G. 2011. AFRA: Argumentation framework with recursive attacks. *Int. J. Approx. Reasoning* 52(1):19–37.
- Baumann, R.; Dvořák, W.; Linsbichler, T.; Strass, H.; and Woltran, S. 2014. Compact argumentation frameworks. In *Proc. ECAI*, volume 263 of *FAIA*, 69–74.
- Brewka, G., and Woltran, S. 2010. Abstract Dialectical Frameworks. In *Proc. KR*, 102–111.
- Brewka, G.; Ellmauthaler, S.; Strass, H.; Wallner, J. P.; and Woltran, S. 2013. Abstract Dialectical Frameworks Revisited. In *Proc. IJCAI*, 803–809.
- Brewka, G.; Polberg, S.; and Woltran, S. 2014. Generalizations of Dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems* 29(1):30–38.
- Caminada, M., and Gabbay, D. 2009. A logical account of formal argumentation. *Studia Logica* 93(2-3):109–145.
- Cayrol, C., and Lagasque-Schiex, M. 2005. On the acceptability of arguments in bipolar argumentation frameworks. In *Proc. ECSQARU*, volume 3571 of *LNCS*, 378–389.
- Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2):321–357.
- Dunne, P. E.; Dvořák, W.; Linsbichler, T.; and Woltran, S. 2013. Characteristics of multiple viewpoints in abstract argumentation. In *Proc. DKB*, 16–30.
- Dunne, P. E.; Dvořák, W.; Linsbichler, T.; and Woltran, S. 2015. Characteristics of multiple viewpoints in abstract argumentation. *Artif. Intell.* 228:153–178.
- Dyrkolbotn, S. K. 2014. How to Argue for Anything: Enforcing Arbitrary Sets of Labellings using AFs. In *Proc. KR*, 626–629.
- Egly, U.; Gaggl, S. A.; and Woltran, S. 2010. Answer-set programming encodings for argumentation frameworks. *Argument & Computation* 1(2):147–177.
- Ellmauthaler, S., and Strass, H. 2014. The DIAMOND system for computing with abstract dialectical frameworks. In *Proc. COMMA*, volume 266 of *FAIA*, 233–240.
- Gaggl, S. A.; Manthey, N.; Ronca, A.; Wallner, J. P.; and Woltran, S. 2015. Improved answer-set programming encodings for abstract argumentation. *TPLP* 15(4-5):434–448.
- Gebser, M.; Kaminski, R.; Kaufmann, B.; and Schaub, T. 2012. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers.
- Gogic, G.; Kautz, H.; Papadimitriou, C.; and Selman, B. 1995. The comparative linguistics of knowledge representation. In *Proc. IJCAI*, 862–869.
- Gomes, C. P.; Kautz, H. A.; Sabharwal, A.; and Selman, B. 2008. Satisfiability Solvers. In *Handbook of Knowledge Representation*, volume 3 of *Foundations of AI*. Elsevier. 89–134.
- Linsbichler, T.; Spanring, C.; and Woltran, S. 2015. The hidden power of abstract argumentation semantics. In *Proc. TAFE*, volume 9524 of *LNCS*, 146–162.
- Marek, V. W., and Truszczyński, M. 1999. Stable models and an alternative logic programming paradigm. In *In The Logic Programming Paradigm: a 25-Year Perspective*. Springer. 375–398.
- Modgil, S. 2009. Reasoning about preferences in argumentation frameworks. *Artif. Intell.* 173(9–10):901–934.
- Nielsen, S. H., and Parsons, S. 2006. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proc. ArgMAS*, volume 4766 of *LNCS*, 54–73.
- Niemelä, I. 1999. Logic programs with stable model semantics as a constraint programming paradigm. *Ann. Math. Artif. Intell.* 25(3-4):241–273.
- Polberg, S. 2016. *Developing and Extending the Abstract Dialectical Framework*. Ph.D. Dissertation, TU Wien, Austria. Upcoming.
- Pührer, J. 2015. Realizability of Three-Valued Semantics for Abstract Dialectical Frameworks. In *Proc. IJCAI*, 3171–3177.
- Strass, H., and Wallner, J. P. 2015. Analyzing the Computational Complexity of Abstract Dialectical Frameworks via Approximation Fixpoint Theory. *Artif. Intell.* 226:34–74.
- Strass, H. 2015. Expressiveness of Two-Valued Semantics for Abstract Dialectical Frameworks. *J. Artif. Intell. Res. (JAIR)* 54:193–231.